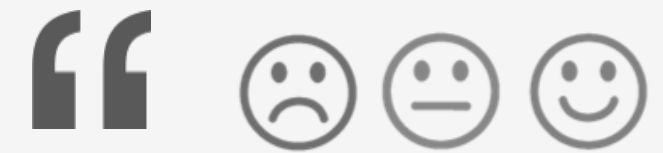


온라인 커뮤니티 특화 감성 사전 구축을 위한 신조어의 극성값 분석 시스템

빅데이터미네이터



주제 :

온라인 커뮤니티 특화 감성 사전 구축을 위한
신조어의 극성값 분석 시스템 ””

<지난주 주요 진행사항>

- 교수님 피드백 반영하여 수정 후, 멀티미디어 학회 논문 최종 제출
- 특허 아이디어 빌딩

진행 사항 - 특허 아이디어 빌딩

💡 KEY POINT

⇒ 의미적 유사도 & 음소 단위 편집거리 (Edit Distance) 알고리즘 함께 적용

- 1) 형태변형어가 입력되면 word2vec을 사용하여 의미적 유사도가 높은 후보 단어를 5개 추출한다.
- 2) 형태변형어와 후보 단어를 각각 쌍으로 만든다. [형태변형어, 후보어] 5 쌍 생성.
- 3) 각 쌍에 대해 편집거리 알고리즘을 적용하여, 한 글자씩 음소 단위 편집거리를 계산한다.
- 4) 각 후보 단어별 음소 편집거리의 합계를 구한다.
- 5) 편집거리 합계 중에서 가장 0(원본 = 0)에 가까운 후보어가 형태유사어로 선택되어 사용자에게 편집거리 수치와 함께 제공된다.

진행 사항 - 특허 아이디어 빌딩



1) 형태변형어가 입력되면 word2vec을 사용하여 의미적 유사도가 높은 후보 단어를 5개 추출한다.

(예시)

→ 형태변형어 ‘댕댕이’ 입력

→ 후보 단어 ‘멍멍이’, ‘강아지풀’, ‘강아지’, ‘장난감’, ‘개’ 추출

2) 형태변형어와 후보 단어를 각각 쌍으로 만든다. [형태변형어, 후보어]의 쌍이 총 5개 생성된다.

(예시)

[댕댕이, 멍멍이]

[댕댕이, 강아지]

[댕댕이, 장난감]

[댕댕이, 강아지풀]

[댕댕이, 개]

진행 사항 - 특허 아이디어 발딩



3) 각 쌍에 대해 편집거리 알고리즘을 적용하여, 한 글자씩 음소 단위 편집거리를 계산한다.

→ 단 편집거리 알고리즘에서 str 길이가 다른 경우, 공백에 대한 삽입/삭제이므로 형태적 차이로 간주하여 음소 단위 편집거리를 1로 설정한다. 따라서 음소 비교를 하지 않고 바로 편집거리를 1로 반영한다.

Ex) [댕댕이, 멍멍이]

댕:멍 → ㄷ/ㅁ, ㅍ/ㅍ ⇒ 2

댕:멍 → ㄷ/ㅁ, ㅍ/ㅍ ⇒ 2

이:이 ⇒ 0

형태변형어	댕댕이	댕댕이	댕댕이	댕댕이	댕댕이
후보 단어	멍멍이	강아지풀	강아지	장난감	개
편집거리	2 + 2 + 0	2 + 3 + 1 + 1	2 + 3 + 1	2 + 3 + 3	2 + 1 + 1
편집거리 합계	4	7	6	8	4

진행 사항 - 특허 아이디어 빌딩



4) 각 후보 단어별 음소 편집거리의 합계를 구한다.

형태변형어	댕댕이	댕댕이	댕댕이	댕댕이	댕댕이
후보 단어	멍멍이	강아지풀	강아지	장난감	개
편집거리	$2 + 2 + 0$	$2 + 3 + 1 + 1$	$2 + 3 + 1$	$2 + 3 + 3$	$2 + 1 + 1$
편집거리 합계	4	7	6	8	4

5) 편집거리 합계 중, 0에 가장 가까운 후보어가 최종 형태유사어로 선택되어 사용자에게 제공된다.

(0에 가까울수록 원형인 ‘댕댕이’와 형태가 가장 유사한 것으로 판별)

→ 단 합계 점수가 같은 후보어의 경우, 원형과 str길이가 가장 가까운 후보어를 최종 선택

→ 예시에서는 최종적으로 ‘멍멍이’가 형태유사어로 제공

차주 계획

1. 구현 결과물 정리
2. 최종 보고서, PPT, 동영상 제작 (논문 및 드림 패스 제출)
3. 특허 아이디어 정리 및 발전



감사합니다 :)
빅데이터미네이터